

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2003, 08:45:46 ; Search time 90 Seconds

(without alignments)  
1027.946 Million cell updates/sec

Title: US-09-768-781-3

Perfect score: 2316

Sequence: 1 MDRVVEIPEEPNVDPVSSLE.....RTRVENSEPPFTEARQSVV 449

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	939.5	40.6	446	11 Q9QXY7	Q9QXY7 mus musculus
2	152	6.6	395	4 Q9H6D3	Q9H6D3 homo sapien
3	141.5	6.1	505	4 Q96P28	Q96P28 homo sapien
4	138.5	6.0	362	4 Q8TBA0	Q8TBA0 homo sapien
5	130	5.6	362	11 Q8R118	Q8R118 mus musculus
6	129	5.6	439	5 O17386	O17386 caenorhabdi
7	120.5	5.2	382	17 Q9V2C2	Q9V2C2 pyrococcus
8	119.5	5.2	745	10 Q48539	Q48539 oryza sativ
9	119.5	5.2	943	10 Q942C7	Q942C7 oryza sativ
10	119	5.1	223	4 Q9NUG5	Q9NUG5 homo sapien
11	118.5	5.1	783	10 Q9SAK8	Q9SAK8 arabidopsis
12	113.5	4.9	508	16 Q98QC2	Q98QC2 mycoplasma
13	112.5	4.9	424	8 Q9B8Z8	Q9B8Z8 schistosoma
14	112.5	4.9	785	10 Q9SA37	Q9SA37 arabidopsis
15	111.5	4.8	424	8 Q94UX1	Q94UX1 schistosoma
16	109.5	4.7	424	8 Q94UX2	Q94UX2 schistosoma

17	109.5	4.7	424	8 Q94UX6	Q94UX6 schistosoma
18	109.5	4.7	461	16 Q97E21	Q97E21 clostridium
19	109.5	4.7	472	2 Q06039	Q06039 lactococcus
20	109.5	4.7	498	8 Q47543	Q47543 chlamydomon
21	109.5	4.7	1776	5 Q963L7	Q963L7 schistosoma
22	108.5	4.7	424	8 Q94UX8	Q94UX8 schistosoma
23	108	4.7	459	8 Q21706	Q21706 pygathrix b
24	107	4.6	459	8 Q21845	Q21845 pygathrix b
25	107	4.6	459	8 Q21707	Q21707 pygathrix b
26	107	4.6	740	16 Q8ZQX3	Q8ZQX3 salmonella
27	106.5	4.6	341	16 Q9PPX3	Q9PPX3 ureaplasma
28	106	4.6	684	16 Q8XM39	Q8XM39 clostridium
29	105.5	4.6	459	8 Q21708	Q21708 pygathrix b
30	105.5	4.6	459	8 Q955T9	Q955T9 cheirogaleu
31	105.5	4.6	459	8 Q955T7	Q955T7 cheirogaleu
32	105.5	4.6	511	2 Q9XDK3	Q9XDK3 bacteroides
33	105	4.5	740	16 Q8Z874	Q8Z874 salmonella
34	105	4.5	995	5 Q9N3H7	Q9N3H7 caenorhabdi
35	104.5	4.5	741	16 Q8X7V5	Q8X7V5 escherichia
36	104	4.5	960	5 Q960V0	Q960V0 drosophila
37	103.5	4.5	459	8 Q955T8	Q955T8 cheirogaleu
38	103.5	4.5	469	8 Q9B6C8	Q9B6C8 yarrowia li
39	103.5	4.5	503	17 Q9UZC2	Q9UZC2 pyrococcus
40	103.5	4.5	759	16 Q8RB81	Q8RB81 thermoplasma
41	103.5	4.5	858	16 Q99XH1	Q99XH1 streptococc
42	103.5	4.5	2143	10 Q9C8A6	Q9C8A6 arabidopsis
43	102.5	4.4	313	5 Q45426	Q45426 caenorhabdi
44	102.5	4.4	403	2 Q8RJL1	Q8RJL1 vibrio chol
45	102.5	4.4	494	16 Q8XN35	Q8XN35 clostridium

#### ALIGNMENTS

RESULT 1  
Q9QXY7 PRELIMINARY; PRT; 446 AA.  
AC Q9QXY7;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE KX antigen (1810038K19RIK protein).  
GN KXK OR KX OR 1810038K19RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C; TISSUE=SKELETAL MUSCLE;  
RX MEDLINE=20009522; PubMed=10541802;  
RA Collec E., Colin Y., Carbonnet F.; Hattab C., Bertrand O.,  
RA Cartron J.P., Kim C.L.;  
RT "Structure and expression of the mouse homologue of the XK gene";  
RL Immunogenetics 50:16-21(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C; TISSUE=SKELETAL MUSCLE;  
RA Le Van Kim C., Collec E., Colin Y.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner T., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Guinacich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., McMaeter P.,  
RA Nordone P., King B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wuzhah-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
RA Hayaishizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
DR EMBL; AF155511; AAF14527.1; -  
DR EMBL; AK007734; BAB2522.1; -  
DR MGD; MGI:103569; Xkh.  
SQ SEQUENCE 446 AA; 51114 MW; D785FB7B9E28B98B CRC64;

Query Match 40.6%; Score 939.5; DB 11; Length 446;  
Best Local Similarity 43.8%; Pred. No. 4.7e-70;  
Matches 180; Conservative 85; Mismatches 135; Indels 11; Gaps 4;  
QY 33 PPSGILSTELVCGEASALYVRIYKNSYTYTTFSPMPSSIMVQLTFVHRDL 92  
DB 3 PPSVIAVFLVFAETAALYLSSTYSAGDRMMQVLTLLPSLPCALVQPTLLFVHRDL 62  
QY 93 AKOKPLSLFHLIILGVPVIRCLEAMIKYTLTWKKEOEPEYVSLTRKK-MLIDGEEVLIE 151  
DB 63 SRDRPLALLMHLQLGPLYRCEVFCYC---QSDNEEPVSVIKKQMPKDGUSEBE 119  
QY 152 WEVGHISIRTLAMHRNAYKRMQSOIAFLGSPQLTLYLVSLISAEVPLGRVVLVMSVLS 211  
DB 120 KEVGOAEGLITRHSAPRASVIOAFLGSAQLTLQVLTIVLEQNTTGRFCFIMTSLLS 179  
QY 212 VTYGATLCNMLAIQIKYDDYKIRIGPLEVLICITWRTLEITSRLLIILVPSATLKLKAVP 271  
DB 180 IVYGLRCNLAIAIKYDEYEVKVKPLAYCIFLRSFEIATRVIVLFTSVLKIWVA 239  
QY 272 FLVNLFIILPEPWIKFWRGSAQPNNEKNFNRVGTGLVLTIVSLTYAGNFCWSALQ 331  
DB 240 VILVNPSPPLVPIVFWCSGSPENIEKALSRGTIVICFLTLVAGINMFCWSAVQ 299  
QY 332 LRLADRLVDKQNGHGLHYSLVRLVENVIMLVKFPFGVKVLYNCHSLIALQLIITAY 391  
DB 300 LKIDNPILISKSNWYRLIYYMTRFIENSVLILLWYFFKTDIYVYVYCAPLILQLLIGY 359  
QY 392 LISIDFMLFPQYHLPLSLTHNVVD---YLHCVCCHQHPRTVENSEP 438  
DB 360 CTGILFVLVYQFFHPCKGLFSSVSSEFRALLRCACWSS---LARKSSEP 407

RESULT 2  
ID Q9H6D3 PRELIMINARY; PRT; 395 AA.  
AC Q9H6D3;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE CNA: FLJ22371 fis, Clone HRC06680 (Hypothetical 44.7 kDa protein)  
DE (Similar to hypothetical protein FLJ10307).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,  
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.,  
RT "NEDO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE FROM N.A.  
RP Strausberg R.;

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SKIN;  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK026024; BAB15326.1; -  
DR EMBL; BC013379; AAHL3379.1; -  
DR EMBL; BC028564; AAH28564.1; -  
SQ SEQUENCE 395 AA; 44654 MW; 23199BAEEA6964C6 CRC64;  
Query Match 6.6%; Score 152; DB 4; Length 395;  
Best Local Similarity 21.1%; Pred. No. 0.00012;  
Matches 82; Conservative 71; Mismatches 136; Indels 100; Gaps 17;  
QY 75 MFSSIMVQLTLIFVHRDLAKDKPLSLFMHLIILGVPVIRCLEAMIKYTLTWKKEOEPEYV 134  
DB 62 LFSWMLRADPAGLHGSQPPRRCLAL-LHLQLGLYVRCVQLRQGLLVW---QDEP-- 115  
QY 135 SLTRKMLIDGEEVLIEWEVGHISIRTLAMHRNAYKRMQSOIAFLGSPQLTLYLVSLIS 194  
DB 116 -----SEFDLAYA-DFLALD---ISMLRLETFLETPAQTLTLVLAIMLOS 156  
QY 195 AEVPLGRVVLVMSVLSVYTGATLCNMLAIQIKYDDY-----KIRLGPLEVLCIT 244  
DB 157 GRAEYQWVGICTSFLGISWALL-----DYHRAURTCLPSKPLLGSGSVIYF 204  
QY 245 IWRLEITTSRLIILVLFSAATL-KLKAVPFLVNLFIILFEPWIKFWRGSAQ-MPNNEKN 302  
DB 205 LWNLLLPVRLVAVLFSALFSPYVALHFLGLMLVLL---WV--WLQTDMPDPSSSEW 259  
QY 303 PSRVGTLVVLISVILYAGINFCWSALQRLADRLVDKQNGHGLHYSLVRLVENV 362  
DB 260 LYRVTV-----ATILY-----FSWFN-----VAEGTRGRATIHFAFLSIL 298  
QY 363 MVLVFKFPGVKVLYNCHSLIALQLII-----AVLISIDFMLFPQYHLPLSLTHNVV 417  
DB 299 LVATVWTHS-----SWLPSPGLPQLWLPVGCPCFLGLALRLVYTHWLHP----- 343  
QY 418 DYLCVCCHQHPRTVENSEPFFETEARQ 446  
DB 344 ----SCCWKPDPPQVDGARSLLSPGYQ 367

RESULT 3  
ID Q96PZ8 PRELIMINARY; PRT; 505 AA.  
AC Q96PZ8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE KIAA1889 protein (Fragment).  
GN KIAA1889.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRIN;  
RX MEDLINE=21456161; PubMed=11572484;  
RA Nagase T., Kikuno R., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XXI.  
RT The complete sequences of 60 new cDNA clones from brain which code for  
RT large proteins".  
RL DNA Res. 8:179-187(2001).  
DR EMBL; AB067476; BAB67782.1; -  
FT NON TER 1  
SQ SEQUENCE 505 AA; 56911 MW; D764B96BC29B3623 CRC64;  
Query Match 6.1%; Score 141.5; DB 4; Length 505;  
Best Local Similarity 20.6%; Pred. No. 0.0011;

```
Matches 65; Conservative 69; Mismatches 121; Indels 61; Gaps 13;
QY 98 LSLFPHLILGSPVIRCEAMIKYLLTWKCEQEBEPYVSLTRKMLIDGEEVLIEWEVS 157
Db 110 LQSLIHLQGLQWRYPHTI--YLGIRSRGGE-----NDRWRFYKMWYIE 153
QY 158 IRTLAMRNAYKMSQIQAFIGSPQLTQLYVSLISAEVPLGRVLMVFLSVSYTYGAT 217
Db 154 YADVSM-----LHLIATFLESAPQLVLQCIIVQTHSLQALQGFATAASLSVLAW-AL 205
QY 218 LCNMLAIQIKYDDYKIRLGPVLCITII--WRTLETSLRLILVFSATLKLKAVPFLV 274
Db 206 ASYOKALDRSDRDKK-----PISYMAVLIQFCWHPFTTAARVITTFALFASVQLYFGIFIV 261
QY 275 LNFILILFEPWIKFMRSGAQMPNNIEKNF--SRVGTLLVLISVTILYAGINFCWSALQL 332
Db 262 LHWCIWTF--WIV-----HCETEFCITKWEIEVFMVVGIIYI--FSWFN----- 302
QY 333 RLADRLVDKQNGHMLHYSVRLVENIMLVLPKFGVGVKVLNLYCHSLIALQLII-AY 391
Db 303 -----VKEGRTRCRLFIYVYVILLENTALSALWLYLKAPQIAD-AFAIPALCVVFSS 354
QY 392 LISIDFLLFPQYLHP 407
Db 355 LTGVFPLMYTAPFHP 370

RESULT 4
QSTBAO
ID Q8TBAO; PRELIMINARY; PRT; 362 AA.
AC Q8TBAO;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to KIAA1889 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LYMPH;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024146; AAH24146.1; -.
SQ SEQUENCE 362 AA; 41672 MW; C19D2D71A2377610 CRC64;

Query Match 6.0%; Score 138.5; DB 4; Length 362;
Best Local Similarity 24.7%; Pred. No. 0.0014;
Matches 58; Conservative 43; Mismatches 106; Indels 28; Gaps 8;

QY 174 IQAFIGSPQLTQLYVSLISAEVPLGRVLMVFLSVSYTYGATLCNMLAIQIKYDDYKI 233
Db 14 LETFLESAPQLVLQYIMLKNSAETLPCVSSVTSLSLAWVLASYHKL-LRDSRDDKK- 71
QY 234 RLGPLEVLCITITWRTLETSLRLILVFSATLKLKAVPFLVFLNFIILFEPWIKFMRSGA 293
Db 72 SMSYRGALHLFWRLFTISSRVISFALFASIFQLYFGIFVHVHWCAMAF--WIIHGTD 129
QY 294 QMPNNIEKNFSRVGTLLVLISVTILYAGINFCWSALQLRLADRLVDKQNGHMLHY 353
Db 130 CMSKWEIILFNMV-----VGIVY-----IFCFWN-----VKEGRTRYRMPAY 167
QY 354 SVRLVENIMLVLPKFP-----FGVKVLNLYCHSLIALQLIIAYLISIDFLLFPQYL 407
Db 168 TIVLTENAALTFLWYPRDPTTDSYAVPALC-VFISFVAGIAMLLYYGVLP 221

RESULT 5
Q8RII8
ID Q8RII8 PRELIMINARY; PRT; 362 AA.
AC Q8RII8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
```

```
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to KIAA1889 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024502; AAH24502.1; -.
SQ SEQUENCE 362 AA; 41643 MW; 51D0F86C5E548017 CRC64;

Query Match 5.6%; Score 130; DB 11; Length 362;
Best Local Similarity 23.8%; Pred. No. 0.0071;
Matches 58; Conservative 42; Mismatches 100; Indels 44; Gaps 9;

QY 174 IQAFIGSPQLTQLYVSLISAEVPLGRVLMVFLSVSYTYGATLCNMLAIQIKYDDYKI 233
Db 14 LETFLESAPQLVLQYIMLKNSAETLPCVSSVTSLSLAWVLASYHKL-LRDSRDDKK- 71
QY 234 RLGPLEVLCITITWRTLETSLRLILVFSATLKLKAVPFLVFLNFIILFEPWIKFMRSGA 293
Db 72 SMSYRGALHLFWRLFTISSRVISFALFASIFQLYFGIFVHVHWCAMAF--WIIHGTD 129
QY 294 QMPNNIEKNFSRVGTLLVLISVTILYAGINFCWSALQLRLADRLVDKQNGHMLHY 353
Db 130 CMSKWEIILFNMV-----VGIVY-----IFCFWN-----VKEGRTRYRMPAY 167
QY 354 SVRLVENIMLVLPKFP-----FGVKVLNLYCHSLIALQLIIAYLISIDFLLFPQYL 404
Db 168 TIVLTENAALTFLWYPRDPTTDSYAVPALC-C-----VFVSFVAGITMLLYGV 218

QY 405 LHPM 408
Db 219 LHPM 222

RESULT 6
O17386
ID O17386 PRELIMINARY; PRT; 439 AA.
AC O17386;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 51.1 kDa protein.
GN F08F1.5
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Fulton B.;
RT "The sequence of C. elegans cosmid F08F1.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
```

DR EMBL; AF026213; AAB71305.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 439 AA; E42AC8F17CD2877E CRC64;  
  
Query Match 5.6%; Score 129; DB 5; Length 439;  
Best Local Similarity 19.4%; Pred. No. 0.011;  
Matches 67; Conservative 62; Mismatches 99; Indels 118; Gaps 15;  
  
QY 97 PLSLPMHLIL--LGPVIRCLEAMIKYLTLWKKEEPEPVVSLTRKKMLIDGEEVLEMEV 154  
DB 126 PLSKMWLVICQMGPLFWYKALY-YGMFPRKSSNEN-----TDGEK----- 166  
  
QY 155 GHSIRTLAMHRNAYKMSQ-----IOAFSGVSPOL----- 184  
DB 167 -----RKCFKMWAEARDAATLLRFFFAFLFESAPQLIIQGSIAAASYFONYOTGTG 216  
  
QY 185 TYQLYVSLISAEVPLGRVVLVPSVSVTYGATLCNMLAIQIKYDDYKIRGLPLEVLCIT 244  
DB 217 PYWLYFOAAS-----LLLSIISWSVVQNRSLRMR--DDKVNIPHEAVLQF 264  
  
QY 245 IWRTEITSRLILVFSATLKLKAVPFLVNLFLILPEPWIKFWRSGAQMNNIEKNFS 304  
DB 265 CWRZTLILARIITIVA-----LVLI-----GINVAIHIDA-CT 297  
  
QY 305 RVGTLVVLISVITYLAGINFSCWSALQLRLADRLVDKQNGWGHMGLHYSVRLVENVIMV 364  
DB 298 HIEKLLLIN-TPIHIFPP-----NNVEGNTWRVYL-TAYSEVFIE--WM 339  
  
QY 365 LVKPFQGVKVLNCHSLIALQLIAYLISIDFMLLFFQYHLPLR 409  
DB 340 LVCWLLPLSLNTPPYIEKVQGVPISTFIAGIAMMYQFFHPNR 384  
  
RESULT 7  
Q9V2C2 PRELIMINARY; PRT; 382 AA.  
AC Q9V2C2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein PAB2250.  
GN PAB2250.  
OS Pyrococcus abyssi.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=29292;  
RN SEQUENCE FROM N.A.  
RP STRAIN=ORSAY;  
RA Heilig R.;  
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ248283; CAB49076.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 382 AA; 41459 MW; 08FBA399DAC5AE8C CRC64;  
  
Query Match 5.2%; Score 120.5; DB 17; Length 382;  
Best Local Similarity 22.7%; Pred. No. 0.046;  
Matches 83; Conservative 56; Mismatches 151; Indels 75; Gaps 14;  
  
QY 72 SFFN---PSSIMVOLTILFVHRDLAKDKPLSLFPHLILLLGPVIRCLEAMIKYLTLWKKEE 128  
DB 49 SFLIARAFSSLSGL-----LEKQ-----RLIYLGSVTMAGNALIVHL----- 88  
  
QY 129 QEEPVSVLTKKMLIDGEEV-----IEWEVGHISIRTLAMHRNAYKMSQIQAPLGSVP 182  
DB 89 -----YPLTTSWQVVGKILNGLNGLSWIPAIASASPDNVRARVTSVYFPLASVA 143  
  
QY 183 QLTVQLYVSLISAEVPLGRVVLVPSVSVTYGATLCNMLA---IQIKYDDYKIRGLPLE 239  
DB 144 SVIGN-VVYAQMABLTLKEQWV---VASIFLTLTALSMFLAYLLFNYPVTPKRGKGNVE 198

QY 240 VLCITIMRTLEITSRLILVLFSA-----TLKLKAVPFLVNLFLIL 281  
DB 199 ELHLDPRKVLVITSLIAITAFASGEITYVYVSEALGLGKGTAKLIGWTGFIATAL---- 255  
  
QY 282 FEPWIKFWRSGAQMNNIEKNFSRVGTTLVVLIS-----VTLYAGINFSCWSALQLR-LA 335  
DB 256 --SVYTSWRADVGVKREVMVSLLSMLASPLLSIKTPTIV-FLGIFLALFSAQSFRTS 312  
  
QY 336 DRDLVDKQNGWGHMGLHYSVRLVENVIMVLFKPFQGVKVLNCHSLIALQLIAYLIS 395  
DB 313 RKLVAVYRRSIAIG---GLNAVQNVSTFLGGLFGLAYSIGELHSIITVNLGLASFLEP 369  
  
QY 396 DFMLL 400  
DB 370 SIALI 374  
  
RESULT 8  
O48539 PRELIMINARY; PRT; 745 AA.  
AC O48539;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE RbohAosp (Fragment).  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN SEQUENCE FROM N.A.  
RP Keller T., Damude H.G., Werner D., Doerner P., Dixon R.A., Lamb C.;  
RT "A plant homolog of the neutrophil NADPH oxidase gp91phox subunit gene encodes an intrinsic plasma membrane protein with Ca2+-binding and RANGAP1 domains."  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF015302; AAB87790.1; --  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR002916; Ferric\_reduct.  
DR InterPro; IPR000778; GP91PHOX.  
DR Pfam; PF00036; ehand; 1.  
DR Pfam; PF01794; Ferric\_reduct; 1.  
DR PRINTS; PR00466; GP91PHOX.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_1.  
FT NON\_TER 1  
SQ SEQUENCE 745 AA; 85335 MW; 8734D3E13A46B3B2 CRC64;  
  
Query Match 5.2%; Score 119.5; DB 10; Length 745;  
Best Local Similarity 20.7%; Pred. No. 0.11;  
Matches 89; Conservative 66; Mismatches 171; Indels 103; Gaps 18;  
  
QY 40 STFLYC-GEASALYVMYRIYRNKSETYMTY-TFSFFMFSSIMVOLTILFVHRDLAKDKP 97  
DB 11 SDFAEICGMTESKEFALELPDLSRRQMKVDITNKDLREIWOQIT-----DNSFDSR 64  
  
QY 98 LSLPHLILLLGPVIRCLEAMIKYL-----TLWKKEEPEPVVSLTRKKMLIDGEEV 149  
DB 65 LQIFEMVKNADGRITEAEVKEIIMLSASANKLSRLKEQAEYAAALIMEELDPGLGY 124  
  
QY 150 IEWEVGHISIRTLAMHRNAYKMSQI-----QAFGLSVPOLTVOLYVSLISAEVPLGRVVL 204  
DB 125 ELWQ-----LETLLLOKOTYMYNSQALSVTSQALSNLAGLKKSKIRKIS----- 170  
  
QY 205 MVFSLSVVYTGATLCNMLAIQIKYDDYKIRGLPLEVLCITI-----WRTLEITSRLIL 258  
DB 171 -----TSLSY-----YFEDNWKRLWL-ALWIGIMAGLFTWKFMQYRNRYVD 212  
  
QY 259 VL-FSATLKLKAVPFLVNLFLILFE-----PWIKFWRSGAQMNNIEKNFSRVGTLVV 311  
DB 213 VMGYCVTTAKGAETLKLNMAIILLPVCNRTITWLRSTRAARALPFDNINPHKTIARAI 272  
  
QY 312 LISVTLYAGINFSCWSALQLRLADRLVDKQNGWGHMGLHY-----SVRLVENVIMVLV 366

[illegible]

```

OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vyetotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Kremenetskaia I., Luros J., Lee J.M., Gonzalez A., Altafi H.,
RA Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Huizer L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federopiel N.A., Theologis A.,
RA "Arabidopsis thaliana chromosome 1 BAC T8K14 sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007202; AAD30236.1; -.
DR InterPro; IPR000676; NAD_Exchange.
DR Pfam; PF00999; Na_H_Exchange; 1.
SQ SEQUENCE 783 AA; 88203 MW; 0A6846910C381648 CRC64;

Query Match 5.1%; Score 118.5; DB 10; Length 783;
Best Local Similarity 19.6%; Pred. No. 0.14;
Matches 89; Conservative 69; Mismatches 166; Indels 129; Gaps 19;

QY 73 FFMFSSIMVQLTIPVHRDL-----AKDKPLSLFMHLILGVPVIRCLEAMIKYLTLM 124
DB 16 ENPLNTPIQACILVPSQLFVLLKPGQAGPVAQLAGVLSVLLSRIPKVEFFLQ 75

QY 125 KKEEQEPEYVS--LTKKKMLIDGEEVLEWVGHSIRTLAMHRNAYKMSQI--QAFL-- 178
DB 76 KNAADYSPFPALRSTSPMLIGLEVDLHF-----MERN-FKKAAVITLSPFVVS 124

QY 179 GSVPLTQVLYVLSISAEVPLGRVLMVPSVSVTYGATLCNMLAIQIKYDDYKI---RL 235
DB 125 GLLSFASLMPLPIFGIKEDYPTFFLVLLVLTLSNTASPVVRSIA-----DWKLTNCBI 178

QY 236 GPLEVLCITIVRTLEITSRLILVFSATLKLKAVPFLVNLFIILPE-----PWIKFPRS 291
DB 179 GRLTISCALFIELTNVLYTIMAFISGTIILELFLFLATVALLINNVLAPELW---- 233

QY 292 GAQMPNNIEKNFSRVGLTVLVI-----SVTILYAGINFSCWSALQLRLA 335
DB 234 --PKRNPKEKYSRAETLVFFLILIGITIESYDVSNSVSFAIGINF-----RQKGT 286

QY 336 DRDLVDK-----GQNGHGMGLHYSV-----RLVENVMVLVPKPGV----- 372
DB 287 HRTLQRLSYPIHEPVLVPYGYIGFRFSIIALTFRFYLGIIVIVITAGKFGVIGAS 346

QY 373 -----KVLN--NY-----CHSLIALQLIAYLIS--- 394
DB 347 MYLKIPKKYVLPFLPILSVKGVGLLSDNSYSEKKWMTTTHDMVVAALVITLVSGVL 406

QY 395 IDFMLLFFQ--YLHPLRLSLFTHNVVDYLHCVC 425
DB 407 ASFLKTRKDPAYEKTSLESHTNEELRIISC 439

RESULT 12
Q98QC2 PRELIMINARY; PRT; 508 AA.
AC Q98QC2; 2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein MYPU_4440.
GN MYPU_4440.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
NCBI_TaxID=2107;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=2167165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moser I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;

```

```

RT "The complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis.";
RL Nucleic Acids Res. 29:2145-2153 (2001).
DR EMBL; AL45564; CAC13617.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 508 AA; 61075 MW; 46E82551297E7642 CRC64;

Query Match 4.9%; Score 113.5; DB 16; Length 508;
Best Local Similarity 18.3%; Pred. No. 0.24;
Matches 81; Conservative 90; Mismatches 159; Indels 113; Gaps 20;

QY 28 NRPFTFPFSILFSTLYCGEASALYVMVRIYRKN-----SETVMTYTFSG--FMFS 77
DB 95 NHRNFYIILWFSIYFSG-----IFLFSRNPFTFNEFVELGYIFYFVIYVIFL 147

QY 78 SIMVOLTILFVHRDLAKPKLSLFMHLILGVPVIRCLEAMIKYLTLMKKEQEEPVSLT 137
DB 148 EILYSVYVWLLYNRRVERDHLKAYKLIVFL--IVSMLYVLI--FSSWINKSQ-----IT 197

QY 138 RKKMLID-----GEEV--LIEWEVGHSIRTLAMHRNAYKMSQIQAF 178
DB 198 TDAKLEEWLFPYQNRIOAFVEASDEIVGLIFRIGVVIAYLIL---IPPFPSLLLS 254

QY 179 GSVPLT-----YQLYVLSISAEVPLGRVLMVPSVSVTYGATLCNMLAIQIKYD 229
DB 255 PKKPKRTKVKKKKMLFQNLVSIIVLSTFLNYWILFKQNVVSP-----LEIN 303

QY 230 DYKIRGLPLEVLCIT---IWRTEITSRLILVFSATLKLKAVPFLVNLFIILFEPW 285
DB 304 NLFGLNVVFLVFTFLILWNRRLKSS--IVFMFSVSGF-----FIVVSFTLIIFH-- 353

QY 286 IKFWRSGAQMNPNEKNFSRVGLTVLISVILYAGINFSCWSALQLRLADRLVDKQGN 345
DB 354 ISYDR-----ENKISMFNLIFFSVLTILFILFSNYEMSTLQKQSNFL----- 398

QY 346 WGHMGLHYSVRLVENVMVLVPKPGVKNVLYNCHSLIALQLIAYLISIDFMLLFFQYL 405
DB 399 -----LLTSFSL--YIFMSVTHKLDLIVIANFLDPYLLILLVGFVAVNLFFVFSFY 452

QY 406 HPLRLSLFTH-----NVVDY 419
DB 453 KSLSTLKTEKKPWTFKKQNFIDY 475

RESULT 13
Q9B8Z8 PRELIMINARY; PRT; 424 AA.
ID Q9B8Z8
AC Q9B8Z8; 2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3).
GN NAD4.
OS Schistosoma japonicum (Blood fluke).
OC Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
NCBI_TaxID=6182;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ANHUI;
RX MEDLINE=20349913; PubMed=10899225;
RA Le T.H., Blair D., Agatsuma T., Humair P.F., Campbell N.J.,
RA Iwagami M., Littlewood D.T., Peacock B., Johnston D.A., Bartley J.,
RA Rollinson D., Herniou B.A., Zarlenga D.S., McManus D.P.;
RA "Phylogenies inferred from mitochondrial gene orders-a cautionary tale
RT from the parasitic flatworms.";
RL Mol. Biol. Evol. 17:1123-1125 (2000).
CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR EMBL; AF215860; AAG13138.2; -.
DR InterPro; IPR003918; NADHub_oxred4.
DR InterPro; IPR001750; Oxidored_q1.

```

[illegible]

```

RESULT 14
Q9SA37
ID Q9SA37 PRELIMINARY; PRT; 785 AA.
AC Q9SA37;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE F309.18 protein.
OS F309.18.
GN Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.
OC NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
RA Vysotskaya V.S., Schwartz J., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J., Li J., Kremetska I., Liu A., Luros J., Gonzalez A.,
RA Altafi H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
RA Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Eckert J.R., Federspiel N.A., Theologis A.;
RT "The sequence of BAC F309 from Arabidopsis thaliana chromosome 1.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; ACC06341; AAC34690.1; -.
DR InterPro; IPR00676; NAF Exchng.

```

[illegible]

```

RESULT 15
Q94UX1 PRELIMINARY; PRT; 424 AA.
Q94UX1;
AC Q94UX1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE NADH dehydrogenase subunit 4.
GN NAD4.
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
RN [1]
RP MEDLINE=20349913; PubMed=10889225;
RX Le T.H., Blair D., Agatsuma T., Humair P.F., Campbell N.J.,
RA Iwagami M., Littlewood D.T., Peacock B., Johnson D.A., Bartley J.,
RA Rollinson D., Herniou E.A., Zarlenga D.S., McManus D.P.;
RT "Phylogenies inferred from mitochondrial gene orders-a cautionary tale
RT from the parasitic flatworms.";
RL Mol. Biol. Evol. 17:1123-1125(2000).
RN [2]
RP MEDLINE=20349913; PubMed=10889225;
RX Le T.H., Blair D., McManus D.P.;
RT "Revisiting limited genetic variation within Schistosoma japonicum
RT populations ";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR ENBL; AF412213; AAL2155.1; -.
DR InterPro; IPR001750; Oxidored_g1.
DR Pfam; PF00361; oxidored_g1; 1.

```

KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.  
SQ SEQUENCE 424 AA; 47725 MW; 1BC880B001235E31 CRC64;

Query Match 4.8%; Score 111.5; DB 8; Length 424;  
Best Local Similarity 19.5%; Pred. No. 0.29;  
Matches 90; Conservative 77; Mismatches 156; Indels 139; Gaps 20;

QY 32 TTPRSILFTFLYCGEASALVMVRIYKNSETYRTMTTFSPFMESSIMVOLT-LIFVHR 90  
DB 18 TPIFLIIFYSIWSDSMWVGVIYLCGLVIIDT-----LSCLMIFLTSIIWLVL 70  
QY 91 DLAKDKPLSLFMHLLILGPIRCLEAMIKY-----LTLWKKEEQEPEPVSLTRKQMLIDG 145  
DB 71 WLVGSKDIVLP-----ISVFSAMLYTVVNSLVFWFFYELS-----IISALYMLIVG 117  
QY 146 -----REVLIIEVGHISIRTLAMHRNAYKRMISOIAPLGSVPOLTYQLYVSLISAEVPLGR 201  
DB 118 SPYPERYISSWYFGGYI-----LLSSVPLLIGICFTGLNSGSF---N 156  
QY 202 VTLMVFSLSVSTYGATLCNML-----AIOIKYDDYKIRLGP 237  
DB 157 VILWDKGMCDISYGAFLIIIVNFLTPIVPFPHGWLPLVHAEASSPVSIILSGYIMKIGL 216  
QY 238 LEVLCITITWRTLEITRLLIILVLSATLKLKAVPELVNLFILPEPWIKFWRSGAOMPN 297  
DB 217 VGLVRLCGWLLID-----YIYFSTPLLCYSVVELV-----AAVECDSKRWLAYLSLSH 266  
QY 298 NIEKNFSRVGTLWVL-----ISVTILYAGINFSCWSALQLRLADRDL 339  
DB 267 IL-----IGVCILLTSTVCGDYLAFLVCLGHLSVALLFMII-----WFGYE----- 308  
QY 340 VDKQNGHGMGLHYSVRLVENVIMVLP-----KFFG-VKVLINLY--CHSLIALQ 386  
DB 309 IGSRRNWGILVKIPGGGLIMHPIMGFVFLNVCVGPFPALQFGEELWLVINYITLGDIIISLL 368  
QY 387 LIIAYLIS---IDFMLLPFQVILHPLRSIFTHN--VVDYLHCV 423  
DB 369 LVSIIYIFSGSIIGFIYGLVICSPINTSYEYSGGLDNFLFCI 410

Search completed: April 1, 2003, 08:50:40  
Job time : 96 secs